Inventory of Research Data Management Services in France

Violaine REBOUILLAT[1]

Dicen-IDF, Conservatoire National des Arts et Métiers

Abstract. Data has become more and more ubiquitous in the research context. As a result, a growing number of services are created to analyze, store and share research data. This has induced the Research Data Working Group of the Digital Scientific Library (BSN10) to launch an inventory of French research data management services, funded by the Ministry of Higher Education and Research. The inventory covers all services that are managed by French institutions and infrastructures and dedicated to public research teams from all fields. Sixty services, provided by forty-five structures, have already been identified and analyzed. The paper describes the methodology used to carry out the inventory and analyzes these first results by service type, scope and research field. It also emphasizes the heterogeneous and emergent nature of the inventoried services.

Keywords. research data, data analysis, data sharing, data archiving, data discovery

1. Introduction

The paper provides an overview of research data management services developed in France. It is part of an ongoing study, funded by the Ministry of Higher Education and Research and led by the Research Data Working Group of the Digital Scientific Library[2] (BSN10). The ubiquity of data and the computing capacities to generate, mine and distribute this data increasingly influences research activities. For scientific and economic reasons, shared services are created to make data management easier for research teams. In this context, the ongoing BSN10 study consists in an inventory of French research data management services. By "research data management services", we mean the providing of human and/or technical resources for digital data management in one or more data lifecycle phases (Figure 1). A service is supplied by what we will call a "structure", i.e. by an entity in a research institution or by independent research infrastructure.

The inventory covers all services that are managed by French institutions and infrastructures and dedicated to public research teams from all fields. The aim is to:

- Better know which kind of research data services exist in France and how they are managed;
- Help researchers and librarians to identify structures most able to provide them the appropriate data management support;

---

[1] Corresponding Author, Dicen-IDF, Conservatoire National des Arts et Métiers, 292 Rue Saint-Martin, 75003 Paris, France; E-mail: violaine.rebouillat.auditeur@lecnam.net.

[2] Bibliothèque Scientifique Numérique (BSN) : http://www.bibliothequescientifiquenumerique.fr/

- Identify potential shortcomings and inform political stakeholders about where resources investment is needed.

The paper will first present similarly initiatives of research data services inventories, in France and abroad. It will then describe the methodology used to carry out the inventory and will report the first results. We will conclude by discussing two issues: the sustainability of data management services and their use.

2. State of the Art

Open Science and Open Data movements have increased the interest in research data management services. In 2012, the Royal Society published a report [1], in which research institutions were encouraged to implement data policy and to provide human and technical resources for research data management. Studies were also conducted with the aim to get an overview on existing data management services: Tenopir et al. [2], [3], [4] considered what types of research data services were offered by European and North American academic research libraries; they conducted surveys and showed that libraries offer more commonly informational and consultative services than technical services, such as preparing data for deposit into a repository. Delay-Artous [5] focused on research data services in the humanities and social sciences; she drew a graphical representation of initiatives and stakeholders, while emphasizing how quickly this representation would be obsolete. This point may explain why mappings of research data services also take the form of regularly updated catalogues. The Registry of Research Data Repositories (Re3data) is one of the best internationally known [6], [7]. It focuses on data dissemination and preservation infrastructures. In the Netherlands, the Leiden University has created a catalogue for data management facilities for researchers: the Leiden Research Data Information Sheets [8]. Its scope is larger than the Re3data's, since the catalogue includes not only research data repositories, but also research data archives or tools for data management plan assistance.

3. Methodology

The inventory is based on a four-step methodology.

The first step consisted of identifying structures which provide research data management services. For that purpose, we used different approaches: we investigated conferences on research data topic; we consulted associations of scientific and technical information professionals (i.e. EPRIST [3], Couperin [4] and the URFIST network[5]), because their members are often involved in data management services in their own institutions; sometimes, during meetings, we were also informed by services managers about the existence of other services.

---

[3] EPRIST is the association of scientific and technical information staffs from the French public research organizations: http://www.eprist.fr/.
[4] Couperin is a consortium of higher education and research institutions for access to digital publications: http://www.couperin.org/.
[5] The URFIST network is composed of seven regional scientific and technical information training units: http://urfistinfo.hypotheses.org/.

The second step focused on establishing a typology in which the already identified services were sorted by function. We built on the research data lifecycle (Figure 1).
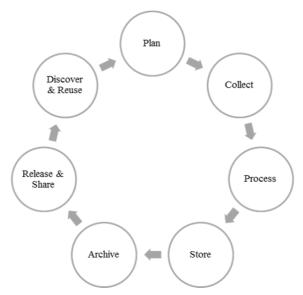
Figure 1. Research data lifecycle.

The typology of services consists of the following categories:

- Information: all websites that aggregate information about news, services, tools or good practices on research data topic;
- Training: face-to-face or distance training services in one or more research data management aspects;
- Support: human resources with IT, documentary, archival and/or legal expertise, which offer to research teams personalized assistance in research data management;
- Data management tool: a tool enabling data traceability such as data management plans (DMP) or persistent identifiers;
- Acquisition platform: an infrastructure providing human and technical resources to support research teams in data collection;
- Computing center: an infrastructure that provides to research teams high-performance computing resources for simulation, modeling and analysis;
- Data registry: an online database that describes scientific datasets;
- Data repository: an online platform that enables users to release or to discover scientific datasets;
- Archiving platform: a platform dedicated to long-term archiving of digital research data.

In the third step, we devised an analysis matrix for each service type, in order to collect information about its identity (name, start date, contact address, etc.), its management (supervisory institutions, human resources, sustainability, etc.), its

functional features (compliance with standards and other technical aspects specific to each service type) and its use (targeted discipline and audience, access conditions, economic model, frequency of use, etc.).

The fourth step of our methodology involved the analysis of the identified services. For each one, we first gathered online documentation (on the website, in papers, reports or communications); then we contacted the service managers for an interview, in order to get additional information. The collected material was recorded in the matrix.

4. First Outcomes

The results reported in this paper should be considered as a photograph of the French data services landscape. They do not match the complete landscape, but only the amount of services inventoried between November 2015 and March 2017. Within that period, 60 services, provided by 45 structures, were actually identified and analyzed. These results are sorted by type in Figure 2.
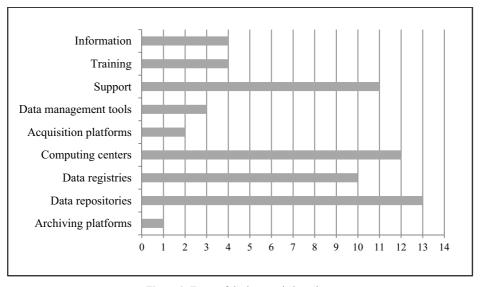

Figure 2. Types of the inventoried services.

Table 1 provides a more detailed insight into results distribution: for each service type, results are sorted both by disciplinary and by geographical scope. Most services (93%) are intended for an institutional or a national audience. Only four services have an international scope: the registry of published astronomical catalogues and tables VizieR and the astronomical objects repository SIMBAD, the sea data repository SEANOE and the archeological data repository ArkeoGIS. Furthermore, institutional services are the most numerous: they represent 63% of the total inventoried services. From a disciplinary perspective and considering the current state of the inventory, French data services seem to cover equally the different research fields: the services in

Human and Social Sciences (HSS) actually number 16; the services in Life and Physical Sciences (LPS) 21; the multidisciplinary services 23.

Table 1. Scopes and research fields of the inventoried services

		Institutional Scope	National Scope	Total
Information	HSS[6]	1	0	1
	LPS[7]	1	2	3
	Multidisciplinary	0	0	0
Training	HSS	2	0	2
	LPS	0	0	0
	Multidisciplinary	1	1	2
Support	HSS	4	0	4
	LPS	2	0	2
	Multidisciplinary	5	0	5
Data Management Tools	HSS	0	0	0
	LPS	1	0	1
	Multidisciplinary	1	1	2
Acquisition Platforms	HSS	0	1	1
	LPS	0	1	1
	Multidisciplinary	0	0	0
Computing Centers	HSS	0	0	0
	LPS	2	0	2
	Multidisciplinary	7	3	10
Data Registries	HSS	2	0	2
	LPS	4	2	6
	Multidisciplinary	1	0	1
Data Repositories	HSS	0	5	5
	LPS	3	0	3
	Multidisciplinary	1	1	2
Archiving Platforms	HSS	0	0	0
	LPS	0	0	0
	Multidisciplinary	0	1	1
	Total	**38**	**18**	**56**

More generally, we can draw two conclusions about today's French landscape of research data management services. The first conclusion is that it is an emergent landscape: 36% of the services we analyzed have been created after 2014; 30% between 2010 and 2014. Second conclusion is the heterogeneous nature of the landscape: variety and diversity prevail, which reflects the proximity of the services with research communities and institutions. This heterogeneity is visible through the

---

[6] HSS: Human & Social Sciences
[7] LPS: Life & Physical Sciences

different types and scopes of services, as described above. It is also visible through the various career profiles of services managers: these may be librarians, archivists, IT staffs or researchers. We noticed that the closer the service is to research teams, the more often this service is managed by researchers or research engineers.

Three examples among the inventoried services may illustrate the observations made above.

The first example is ECOSCOPE Metadata Portal [9], a non-profit data registry of biodiversity research observatories, launched in 2016. ECOSCOPE is a French observation data infrastructure for biodiversity research, maintained by the Foundation for Biodiversity Research (FRB). The ECOSCOPE Metadata Portal was created to make known the data produced by the 200 French observation observatories in the field of biodiversity. The aim is to foster data sharing between these observatories, which currently tend to work as closed circuits, using few external data. The portal also complies with the INSPIRE Directive. It will be harvested by national and international catalogs, such as Géocatalogue[8], the EU BON Portal[9] and GBIF[10]. In March 2017, ECOSCOPE described 52 datasets. The metadata schema is compliant with the Ecological Metadata Language (EML). ECOSCOPE ensures the quality of the metadata recorded by observatories, before editing them. The main challenge of ECOSCOPE is fostering uptake by the French biodiversity community. At this time, only eleven observatories described datasets in the registry. Regarding the consultation rate, ECOSCOPE does not have any tool yet to measure it.

The second example is the CINES archiving platform [10]. The National Computing Center for Higher Education (CINES) provides resources for long-term archiving of digital data to the French research community. It reports directly to the Ministry of Higher Education and Research. Its staffs have an archival and IT expertise. The CINES selects sustainable file formats, applies the persistent identifier system ARK and is compliant with the Dublin Core metadata schema. The archiving quality is certified by the Data Seal of Approval. The main challenge is that the CINES is the only stakeholder in France to provide archiving resources for digital scientific data. As a result, small research projects cannot access the CINES services, as these are subject to charges and give priority to large data volumes.

The last example is PUDC [11], a platform for human and social sciences data, which we classified in the category "support" of the services types. Created in 2011, the platform is based at the University of Caen Basse Normandie and is supported by the research infrastructure PROGEDO. It is managed by a lecturer in sociology and a research engineer specialized in statistics. Their role consists in offering personalized assistance to PhD students and researchers in human and social sciences from the University of Caen Basse Normandie. They help them to reuse quantitative data, especially these from databanks of national and international social surveys. When data do not exist, they provide them methodological advices to generate, analyze and interpret their own data. In 2016, about sixty persons used the PUDC services. The managers however report that researchers do not turn spontaneously to them; extensive communication efforts are necessary. Yet there is a need, especially for support in data analysis. The core challenge of PUDC is thus to keep attracting users, in order to ensure its sustainability.

---

[8] Géocatalogue: http://www.geocatalogue.fr/
[9] EU BON European Biodiversity Portal: http://biodiversity.eubon.eu/
[10] Global Biodiversity Information Facility: http://www.gbif.org/

5. Conclusion

We report here two issues that occurred in the course of the inventory.

The first issue is about the sustainability of data management services. We wondered what financial and human resources services had and, consequently, if these resources were sufficient to guarantee data preservation. We noticed that funding was a major concern. Services commonly have a long-term mission; yet, they often depend on the research funding system, which consists of short-term grants. They are compelled to constantly seek new sources of funding. We tried to determine how many services were sustainable. We considered a service as sustainable, when funded over the long-term and managed by a structure which has a legal entity. It appeared that 37% of the services we inventoried were not sustainable, i.e. more than a third. This situation has an impact on human resources: in most cases, actually, services managers are small teams of two or three staffs. Technical resources can also be restricted, which may impact data preservation. For instance, due to limited funding, the data repository ORTOLANG is only able to transfer one part of its datasets to CINES for archiving.

The second issue concerns the use of the services. Are the services actually used by researchers? Are researchers aware of their existence? How do managers foster the use of their services? During the analysis phase of the inventory, we had difficulty collecting quantitative data on the use of the services. Indeed, the services managers did not necessarily have usage statistics. Most of the time, only those who requested from researchers the creation of a user account to access the service were able to provide us figures. Moreover, research teams are generally not aware of the existence of data management services. A shift is still visible between data processing services and data releasing services: acquisition platforms, computing centers and archiving platforms are relatively well integrated in research practices, whereas services related to open science, such as data repositories, are less used, probably because they are not currently at the heart of researchers' concerns. Journal articles and other forms of publication are more valuable than research data in nowadays evaluation system.

Both issues are decisive for the future of research data management services. If governments and research funders keep fostering open science, financial resources will be invested in research data services and research teams will be encouraged to use these services. But without political incentives, it may be more difficult for services to gain legitimacy.

Acknowledgements

References

[1]   The Royal Society. *Science as an Open Enterprise*. The Royal Society Science Policy Centre, 2012. Available from: https://royalsociety.org/~/media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

[2] C. Tenopir, B. Birch, S. Allard. *Academic Libraries and Research Data Services: Current Practices and Plans for the Future*, Association of College & Research Libraries, 2012. Available from: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

[3] C. Tenopir, R. Sandusky, S. Allard, B. Birch. Research Data Management Services in Academic Research Libraries and Perceptions of Librarians, *Library & Information Science Research* **36**, Issue 2 (2014), 84–90. Available from: http://www.sciencedirect.com/science/article/pii/S0740818814000255

[4] C. Tenopir, S. Talja, W. Horstmann, E. Late, D. Hughes, D. Pollock, B. Schmidt, L. Baird, R. Sandusky, S. Allard, Research Data Services in European Academic Research Libraries, *LIBER Quarterly* **27**, Issue 1 (2017). Available from: http://www.liberquarterly.eu/articles/10.18352/lq.10180/

[5] C. Delay-Artous, Où sont les données de la recherche ? Essai de cartographie, *Systèmes d'organisation des connaissances et humanités numériques*. ISTE Editions, London, 2017. Available from: https://halshs.archives-ouvertes.fr/halshs-01369745

[6] Registry of Research Data Repositories (re3data.org). Available from: http://www.re3data.org/

[7] M. Kindling, H. Pampel, S. Sandt, J. Rücknagel, P. Vierkant, G Kloska, M. Witt, P. Schirmbacher, R. Bertelmann, F. Scholze, The Landscape of Research Data Repositories in 2015: A re3data Analysis, *D-Lib Magazine*.**23**, Issue 3/4 (2017). Available from: https://doi.org/10.1045/march2017-kindling

[8] Leiden Research Data Information Sheets. Available from: https://vre.leidenuniv.nl/vre/lrd/Pages/information-Sheets.aspx

[9] ECOSCOPE Metadata Portal of Biodiversity Research Observatories. Available from: http://ecoscope.fondationbiodiversite.fr/ecoscope-portal

[10] CINES Archiving Platform. Available from: https://www.cines.fr/archivage/

[11] Plateforme Universitaire de Données de Caen (PUDC). Available from: http://www.unicaen.fr/recherche/mrsh/pudc